DOCUMENT RESUME

ED 416 208 TM 028 055

AUTHOR Soderstrom, Irina R.; Leitner, Dennis W.

TITLE The Effects of Base Rate, Selection Ratio, Sample Size, and Reliability of Predictors on Predictive Efficiency Indices Associated with Logistic Regression Models.

PUB DATE 1997-10-00

NOTE 25p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Chicago, IL, October 15-18, 1997).

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Mathematical Models; Monte Carlo Methods; *Prediction; *Regression (Statistics); Reliability; *Sample Size; *Selection; Simulation

IDENTIFIERS *Base Rate Information; *Logistic Regression

ABSTRACT

While it is imperative that attempts be made to assess the predictive accuracy of any prediction model, traditional measures of predictive accuracy have been criticized as suffering from "the base rate problem." The base rate refers to the relative frequency of occurrence of the event being studied in the population of interest, and the problem stems from the fact that statistical prediction models often are not valid when applied to populations with a different base rate than the population for which the prediction model was constructed. This study tested alternative predictive accuracy models, two of which account for base rate levels, to determine the degree to which they are base rate invariant. The indices were the three indices recommended by S. Menard (1995), the Relative Improvement over Change (RIOC) method, and the percentage correct classification indices. A Monte Carlo simulation study was undertaken to generate two types of logistic regression models, one with a dichotomous predictor and a continuously measured predictor and the other with two dichotomous predictors. Four reliabilities, three base rate conditions, and two sample sizes were used. All three of Menard's indices were found to be sensitive to fluctuations in the base rate. Conditions under which these indices and the RIOC may be used are summarized in a table. It is recommended that researchers compute all three of Menard's indices and then compare the values across the three to get an indication of the underlying base rate of the sample. (Contains 3 tables and 26 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 416 208

The Effects Of Base Rate, Selection Ratio, Sample Size, And Reliability Of Predictors On Predictive Efficiency Indices Associated With Logistic Regression Models

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

IRINA R. SODERSTROM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- ☐ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- • Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Irina R. Soderstrom
Eastern Kentucky University

Dennis W. Leitner
Southern Illinois University at Carbondale

Paper Presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago, IL, October 15-18, 1997

TM028055

BEST COPY AVAILABLE

The central idea of statistical prediction is that previously observed relations between predictor variables and criterion classifications permit estimates of the most probable criterion outcomes for each category of persons or groups. Predictive classifications have many uses for research, program planning and evaluation, policy development, and individual case decision making (Gottfredson & Tonry, 1987). This is particularly true in the medical sciences, and is becoming increasingly popular in the behavioral and social sciences. In fact, prediction models have become central both to setting general policies as well as to making decisions about individuals in a number of behavioral science domains. For example, a search of the ERIC database indicated that between January, 1990 and July, 1997, a keyword search on the term "logistic regression" produced 84 abstracts. A sample of the abstracts which were directly relevant to educational research applications included several studies which employed logistic regression to predict student retention in college (e.g., see Huesman, Moore, Huang, & Guo, 1996; Sherry & Sherry, 1996; Miller, Brownell, & Smith, 1995; Wilson & Hardgrave, 1995; Gillespie & Noble, 1992), to detect Differential Item Functioning of test items (e.g., see French & Miller, 1996; Ryan & Chiu, 1996; Ryan & Bachman, 1992), to predict successful performance of students on various standardized achievement tests (e.g., see Berends, Koretz, & Harris, 1995; Weimer, 1996), and to predict faculty retention and tenure outcomes (e.g., see Eimers, 1995).

Thus, it is imperative that attempts be made to assess the predictive accuracy of any prediction model. Unfortunately, traditional measures of predictive accuracy have all been criticized as suffering from the "base rate problem." The base rate refers to the relative frequency of occurrence (i.e., ratio of successes to failures) of the event being studied in the population of interest. The problem stems from the fact that statistical prediction models often are not valid when applied to populations with a different base rate than the population for which the prediction model was constructed. This problem is further compounded by the fact that most measures of predictive accuracy are highly sensitive to changes in the base rate (Fergusson et al., 1977).

The issues of predictive accuracy and the base rate problem are particularly relevant for logistic regression prediction models where the dependent or criterion variable typically is a dichotomous one. The reason for this is that the base rate for a dichotomous variable becomes the marginal distribution of the outcome expectancy table (and computation of the predictive accuracy index is based on this expectancy table). For example, if 40% of students successfully complete a developmental mathematics program, then a 40% base rate will be used in the logistic regression prediction model. But if that same model is then used on a group of developmental mathematics students representing a population with a true base rate of 20%, then many errors in classification will occur since the prediction model will be invalid for the latter group of students.

In the ERIC-indexed articles listed above, the base rate problem is relevant if someone wishes to replicate those studies. It is impossible to know what the true base rate is for the sample being studied, and this ambiguity causes uncertainty as to the validity in generalizing results across studies. For example, if a researcher wanted to see if one of the student-retention logistic regression prediction models would perform well for another group of students, there would be a problem with directly comparing the predictive efficiency or accuracy indices of those models. The problem would stem from the fact that it would be unclear as to whether or not the two groups of samples originated from populations with identical base rates.

It is the intention of this study to test alternate predictive accuracy indices, two of which account for base rate levels, in order to determine to what degree they are base rate invariant (i.e., the value of the measure is independent of the actual sampling ratio of successes to failures) or less sensitive to base rate changes, as model conditions including selection ratio, reliability levels of the predictor variables, and sample size are varied across two types of logistic regression models.

Significance of the Study

In many disciplines (e.g., medical and health research), logistic regression has become the standard method of analysis for explaining the relationship between explanatory variables and a dichotomous, or binary, response variable (Hosmer & Lemeshow, 1989). The base rate problem and its effects on predictive accuracy indices for logistic regression prediction models is an area of methodological development that has received very little attention, especially in comparison to the methodological development in the area of goodness-of-fit indices. However, Menard (1995) illustrates how a prediction model which fits the data well, can still lead to errors in classification. Even with these errors in classification, Gottfredson & Tonry (1987) posit that "in virtually every decision-making situation for which the issue has been studied, it has been found that statistically developed prediction devices outperform human judgments" (p.36). Thus, it is necessary to advance the methodological development of predictive accuracy indices. And since the biggest problem associated with their use concerns the base rate issue, measures considered to be less sensitive to the base rate, or base rate invariant, need to be assessed to determine under which conditions of the logistic regression model they withstand base rate fluctuations.

Methods

A Monte Carlo simulation study was undertaken to generate two types of logistic regression models. One model had a dichotomous predictor and a continuously measured predictor, while the other model had two dichotomous predictors. All possible combinations (4) of reliability levels in the two predictors were simulated (high-high; low-low; high-low; and low-high). Three base rate conditions were simulated (.10, .30, .50), while a full spectrum of possible selection ratios were employed under both small (N = 200) and large (N = 2,000) sample scenarios.

Controlling Reliability Levels of Predictor Variables

The reliability levels of the continuous predictors were controlled by generating the variables from normal distributions with either low reliability (.60) or high reliability (.90), based on the classical true score definition of reliability (i.e., observed score = true score + error). Specifically, reliability was defined as the proportion of the observed score variance that was true score variance (Allen & Yen, 1979). The two reliability levels were chosen in order to see if differences in the effectiveness of the predictive efficiency indices could be detected when variables with unacceptably low measurement reliability levels (e.g., .60) were employed, as opposed to variables with measurement reliability levels considered to be acceptable (i.e., .90) for use in prediction instruments that may be utilized to make decisions affecting someone's life (Nunnally, 1978).

Low (kappa = .30) and high (kappa = .80) reliability levels were used in the computations of dichotomous explanatory variables as well, but were defined by Cohen's (1960) kappa. Kappa can be interpreted as the amount of agreement-above-chance as a proportion of the maximum possible agreement-above-chance (Collis, 1985).

The decision as to which values were used to represent low and high reliability in the dichotomous predictors was based primarily on guidelines provided by Landis and Koch (1977). The authors described the strength of agreement for a kappa of .30 as "Fair," and for a kappa of .80 as "Substantial" and bordering on "Almost Perfect" (p. 165). Further, Landis and Koch (1977) illustrated that reliability levels for categorical data do not need to be as high as the levels required for continuous data in order to obtain high reliability. They reported that agreement between experts in clinical psychiatric research tends to result in kappa values between .50 and .59, which was interpreted as "Moderate" agreement. However, a kappa of .61 was all that was required to obtain "Substantial" agreement.

Additional information that aided in the choice of a low value of kappa stemmed from a guideline offered by Waltz, Strickland, and Lenz (1991) which said, "An acceptable level of

interrater agreement varies from situation to situation. However, safe guidelines for acceptable levels are P_o values greater than or equal to .80 or k greater than or equal to .25" (p. 242). Thus, it was decided that a kappa of .30 would adequately indicate low reliability for dichotomous predictors, while remaining at an acceptable level that would represent hypothetical research scenarios.

Generating the Logistic Regression Models

The variables were submitted to a logistic regression analysis using the PROC LOGISTIC DESCENDING command in SAS Logistic Regression Examples manual, Version 6 (SAS Institute, 1995), which instructed the program to model predicted probabilities for Y = 1. This program was used to generate the actual logistic regression models, while manipulating the reliability levels (similar to methods used in a study by Soderstrom & Han, 1993) of the explanatory variables. The logistic regression algorithms were based on the maximum likelihood iterative estimation procedure.

Specifically, the data simulation was conducted such that two different forms of the two-predictor logistic regression model were generated: 1) a model comprised of one dichotomous predictor and one continuously measured predictor; and 2) a model comprised of two dichotomous predictors. In all cases, a dichotomous dependent variable was employed.

Therefore, both of the logistic regression models were of the following form (SAS Institute, 1995):

$$\text{logit}(p) = \log(p/(1-p)) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

where: $p = \text{Prob}(Y = Y_1 \mid X_1, X_2)$ is the response probability to be modeled, and Y_1 is the first ordered level of Y.

α is the intercept parameter.

β is the vector of slope parameters.

BEST COPY AVAILABLE

X_1	is the vector of a continuous explanatory variable in the first logistic regression model, and is the vector of a dichotomous explanatory variable in the second logistic regression model.
X_2	is the vector of a dichotomous explanatory variable.

Specifically, the two logistic regression models were simulated using the respective low measurement reliability levels on all variables, within each level of base rate. Next, the two logistic regression models were simulated using the respective high reliability levels on all variables, within each level of base rate. The simulations were then repeated for the two logistic regression models using the respective high reliability values on the first predictor in the model, while employing the respective low reliability values on the second predictor. Finally, the simulations were repeated for the two logistic regression models using the respective low reliability values on the first predictor, while employing the respective high reliability values on the second predictor.

Generating 2 x 2 Classification Prediction Tables

Once 2,000 replications of a given logistic regression model were simulated, 2 X 2 classification tables were generated from the logistic regression results. One marginal distribution of the table was defined by the base rate (i.e., the actual ratio of Y = 1 to Y = 0). The other marginal distribution, the selection ratio, was defined by setting a predicted probability cutoff point to determine who was predicted to have Y = 1 and who was predicted to have Y = 0. This process was repeated across five cutoff point choices (and their associated selection ratios): .10, .30, .50, .70, and .90. Thus, a cutoff point of .10 would mean that all subjects with predicted probabilities for Y = 1 that were greater than or equal to .10 would be selected for that category by the model (a large selection ratio). Conversely, a cutoff point of .90 would result in a very small selection ratio.

While the same cutoff points were applied to all simulated logistic regression models, the resulting selection ratios were dependent upon the predicted probability distribution generated by each individual prediction model. The shape of the predicted probability distribution was influenced by the choice of base rate. Therefore, a normally distributed predicted probability distribution would result in a larger selection ratio at a high cutoff point, than would a highly skewed predicted probability distribution after employing that same cutoff point. The actual values of the selection ratios for each of the simulated logistic regression models were not observed. The primary point of manipulating the cutoff point was to show how the indices fluctuated as the selection ratio got smaller or larger, within a given base rate. It was not an intention of this study to specify what those exact selection ratio levels were.

Computing Predictive Efficiency Indices

Once 2 X 2 classification tables (cross tabulations of predicted success/failure outcomes with observed success/failure outcomes) were generated for all 2,000 replications of a given logistic regression model, three predictive efficiency and accuracy indices (λ_p, τ_p, and ϕ_p) proposed by Menard (1995) for use with logistic regression models, were computed for each corresponding classification table. Two additional indices of predictive accuracy and efficiency, the Relative Improvement Over Chance (RIOC) index and the percentage correct classification, were generated as well. Next, means and standard deviations for each index were computed across the 2,000 replications. These summary data were then tabled and graphed. Thus, in all, 120 logistic regression models were designed and generated through a SAS computer simulation program.

Menard suggested that his adaptations of the three predictive efficiency indices he proposes for use with logistic regression models (i.e., λ_p, τ_p, and ϕ_p) are more appropriate for use with logistic regression models (particularly since two of the three indices account for the observed base rate), yet he failed to demonstrate the efficacy of these measures as various model conditions and base

rates change. This current study addresses a number of research questions pertaining to how these indices of predictive efficiency would perform under specified logistic regression model constraints (see also Soderstrom, 1997). Specifically, these research questions inquired as to how Menard's (1995) three recommended indices of predictive efficiency (i.e., λ_p, τ_p, and ϕ_p), as well as the frequently used Relative Improvement Over Chance (RIOC) and the percentage correct classification indices, would be influenced by fluctuations in the base rate, given various constraints (i.e., changes in measurement reliability levels of predictor variables, changes in sample size, and changes in selection ratio) of the logistic regression model.

Investigation of Sample Size Influence

Additionally, influences of sample size were investigated. Once all 120 logistic regression models were simulated using a large sample size of 2,000 (again, see Tables 3 and 4), a subsample of these models were simulated again using a small sample size of 200. Due to the extensive amount of computer time involved in computing 2,000 replications of each logistic regression model, it was expected that the sample size issue could be adequately addressed by comparing predictive efficiency index variation for only a subsample (16) of the models, representing a cross-section of reliability, base rate, and cutoff point combinations.

Thus, in all, 136 logistic regression models were designed and generated through the computer simulation program. Simulations were replicated 2,000 times for each model so that results could be averaged over the replications. This produced standard errors for the proportions of about .01.

It also should be noted that this study is an extension of earlier investigations of the base rate which did not explore the effects of reliability of the predictor variables on resulting predictive efficiency indices (see Soderstrom & Leitner, 1996). It is expected that this study will fill a methodological void concerning the use of predictive efficiency indices; namely, to indicate under

what types of logistic regression conditions they appear to be less sensitive to changes in the base rate, as the conditions of selection ratio, reliability levels of predictor variables, and sample size are varied.

Results

Because a total of 17 tables and 41 figures were required to present the results of this study in tabular and graphical form, only verbal summaries of these results are presented here. The reader is encouraged to obtain a copy of the dissertation which contains the more detailed results of the dissertation (see Soderstrom, 1997).

The results of this study will be discussed separately as they pertain to each individual efficiency index, to the base rate problem, to the selection ratio, to the measurement reliability of predictor variables, and to the sample size. Following this discussion, tables summarizing the primary findings of this study will be presented.

Φ_p

The results from analyses of the simulated logistic regression models led to the conclusion that ϕ_p tended to yield the highest estimates of predictive efficiency, regardless of base rate or selection ratio. Additionally, ϕ_p was found to be the least sensitive to base rate changes of the three predictive efficiency indices proposed by Menard (1995) for use with logistic regression models.

The results of the analyses performed on simulated logistic regression models led to the conclusion that ϕ_p tended to yield the highest estimates of predictive efficiency, regardless of base rate or selection ration when the model was comprised of a continuous and a dichotomous predictor. But when the model was comprised of two dichotomous predictors, τ_p became the most base rate invariant, followed by ϕ_p, and then λ_p. Further, the results led to the conclusion that when

the logistic regression model was comprised of one continuous predictor and one dichotomous predictor, Φ_p remained close in value to τ_p, regardless of the reliability levels of the predictor variables. However, when the logistic regression model contained two dichotomous predictors, Φ_p was closer in value to λ_p if the base rate was .10, but was closer in value to τ_p if the base rate was .30. All three efficiency indices were identical in value when the base rate was .50. Once again, these patterns were consistent across all measurement reliability combinations.

$\underline{\lambda_p}$

Results from analyses of simulated logistic regression models led to the conclusion that λ_p consistently yielded the lowest and most variable estimates of predictive efficiency, regardless of sample size, base rate, or selection ratio. Additionally, λ_p was found to be the most sensitive to fluctuations in the base rate. The type of predictors (continuous or dichotomous) and their levels of measurement reliability (high or low) did not alter these findings.

$\underline{\tau_p}$

The results of the analyses performed on simulated logistic regression models led to the conclusion that when the logistic regression model was comprised of one continuous predictor and one dichotomous predictor, τ_p remained close in value to Φ_p, regardless of the reliability levels of the predictor variables. However, when the logistic regression model contained two dichotomous predictors, τ_p was closer in value to λ_p if the base rate was .10, but was closer in value to Φ_p if the base rate was .30. All three efficiency indices were identical in value when the base rate was .50. Once again, these patterns were consistent across all measurement reliability combinations.

BEST COPY AVAILABLE

RIOC

The general conclusion made about RIOC upon completion of the simulated logistic regression model analyses was that the index was very consistent across all base rate and cutoff point (and associated selection ratio) combinations, unless both of the predictors demonstrated low reliability. It did not matter whether or not the model contained a continuous or a dichotomous predictor, nor did it matter which predictor had the high reliability. However, if both predictors displayed low reliability, then the RIOC means were quite variable across all base rate and cutoff point scenarios.

Another finding identified with the RIOC index was the fact that when the logistic regression model contained only dichotomous predictors (as opposed to continuous predictors), and when the base rate was .10 or .30 coupled with a high predicted probability cutoff point (i.e., low selection ratio), the RIOC index was not calculable. This type of base rate/cutoff point combination most likely resulted in a large number of false negatives.

The explanation for this inability to compute the RIOC index was that zero values were occurring in the denominator of the index formula. If zeros had occurred in the numerator of the index formula, the index would have returned a value of zero. Since the denominator of the RIOC index is computed as the frequency of random correct (expected) predictions minus the frequency of maximum correct (possible) predictions, it was concluded that the expected frequencies were equal to the possible frequencies in the cases where the index was incalculable.

Percent Correct Classification

Results from analyses of simulated logistic regression models revealed that when the model was comprised of a continuous predictor and a dichotomous predictor, the percent correct classification index consistently indicated good classification ability of the model across most base rate and cutoff point levels. The only exception to this statement was when both predictors

displayed low reliability, which in turn caused the percent of correct classifications to vary considerably across base rate/selection ratio combinations.

When the logistic regression model was comprised of two dichotomous predictors, the general finding for the percent correct classifications index was that the model generally indicated good classification ability at base rate levels of .10 and .50. But when the base rate was .30, the classification ability of the model dropped considerably at high cutoff points (i.e., low selection ratios). Once again, the model with low reliability in both of the predictors was an exception to this statement, since this particular model classified a high proportion of cases correctly when the base rate was .10, but displayed worse classification ability as the base rate approached .50.

Base Rate

Several key findings were obtained regarding the influence of base rate on predictive efficiency indices. First, all three predictive efficiency indices proposed by Menard (1995) for use with logistic regression models were found to be sensitive to fluctuations in the base rate. Φ_p was determined to be the most base rate invariant index, followed by τ_p, while λ_p displayed the greatest sensitivity to base rate changes.

It was not surprising that ϕ_p was found to be the most base rate invariant of the three predictive efficiency indices recommended for use with logistic regression, since it was the only coefficient that took both the base rate and the selection ratio into account in its computation. However, this finding was contradictory to Menard's (1995) suggestion that τ_p would probably be the most appropriate index to utilize when assessing a model's predictive efficiency.

Second, all three of Menard's (1995) predictive efficiency indices yielded means that were more consistent (with themselves and with each other) across selection ratio levels within a given base rate, the closer the base rate was to .50. This finding led to the conclusion that as long as the base rate was close to .50, it did not matter which index was utilized. But if the base rate was

much less than .50, the estimates of predictive efficiency yielded by the three indices would vary considerably.

Thus, this finding was consistent with Davis' (1971) recommendation which was cited in Smith (1996, p. 94), and which suggested that "researchers seek a 50-50 split in studying associations of dichotomies and avoid dichotomies more extreme than 30:70." Smith concurred with this advice.

The results from analyses of simulated logistic regression models indicated that the model type (i.e., either one continuous and one dichotomous predictor, or two dichotomous predictors), as well as the reliability levels of the predictors, did have additional influence on the predictive efficiency indices.

It was observed that when the model was comprised of one continuous and one dichotomous predictor, the indices were much more base rate invariant than when the model contained two dichotomous predictors. Measurement reliability levels of the predictors did not seem to alter this finding, except for when both predictors displayed low reliability. This condition caused all three efficiency indices to yield much lower estimates of predictive efficiency than was the case for the model containing a continuous and a dichotomous predictor. Yet, even when both predictors displayed low reliability, all three indices became more stable (i.e., closer in value) when the base rate was .50.

Further, when the base rate was .30 or .50, predictive efficiency index means did not vary across cutoff point levels within the given base rate. But when the base rate was .10, the index means even varied across cutoff points.

It also was observed that when the logistic regression model was comprised of two dichotomous predictors, the three predictive efficiency indices were much more sensitive to base rate changes than was the case for the model containing a continuous and a dichotomous predictor. This conclusion derived from the observation that all three of Menard's (1995) indices

tended to be consistent in value across cutoff point levels within any given base rate, but were extremely inconsistent in value across base rate levels for the model comprised of two dichotomous predictors.

Another interesting finding which resulted from analyses of the simulated logistic regression models was that while ϕ_p was the most base rate invariant index for the model comprised of one continuous and one dichotomous predictor, τ_p was the most base rate invariant index for the model comprised of two dichotomous variables. But for either model type, λ_p was always the most sensitive to base rate changes.

The finding that τ_p was more base rate invariant than ϕ_p when the logistic regression model contained two dichotomous predictors was consistent with Menard's (1995) recommendation to select τ_p to assess a logistic regression model's predictive efficiency. In his monograph, Menard (1995) demonstrated logistic regression analyses using only categorical predictor variables. This observation, coupled with the findings of the current study regarding the model comprised of only dichotomous variables, allowed for speculation as to why Menard recommended the use of τ_p over ϕ_p.

Selection Ratio

The conclusion particularly relevant to the selection ratio which was derived from the analyses of the simulated logistic regression models was that regardless of the base rate level, the efficiency indices always indicated improved predictive efficiency as the selection ratio approached the value of the base rate. Additionally, it was concluded that there was much less variation in index means across predicted probability cutoff points within any given base rate level if the model contained only dichotomous predictors. But if a continuous predictor was included in the model, index means varied more across cutoff points within any given base rate.

BEST COPY AVAILABLE

Measurement Reliability Of Predictor Variables

Results from analyses of the simulated logistic regression models led to several conclusions regarding the measurement reliability of the predictor variables. When the model was comprised of one continuous and one dichotomous predictor, the indices were not as sensitive to base rate changes as long as the continuous predictor was measured with high reliability. Thus, when the logistic regression model had either low reliability on both predictors, or low reliability on the continuous predictor but high reliability on the dichotomous predictor, index means were quite variable across base rate and cutoff point levels. On the other hand, if the model at least had high reliability on the continuous predictor, index means were much more consistent across base rate and cutoff point (and their associated selection ratio) levels.

When the logistic regression model was comprised of two dichotomous predictors, the conclusion was slightly different. It was observed that as long as at least one of the two dichotomous predictors displayed high reliability, the influence of base rate changes was similar across reliability combinations (i.e., high reliability on both predictors; high reliability on the first predictor and low reliability on the second predictor; low reliability on the first predictor and high reliability on the second predictor). But for the model with low reliability in both predictors, all base rate and cutoff point combinations yielded index means that were consistently low.

Sample Size

The key conclusion made regarding the influence of sample size was that it had very little impact on the mean values of the indices across the 2,000 replications. However, the standard errors associated with these means were substantially larger at smaller sample sizes. Thus, it was concluded that more variability in computed indices across samples should be expected as sample size decreases.

The results just presented for simulated logistic regression models with a continuous predictor and a dichotomous predictor can be found in table 1. Similarly, the results just presented for simulated logistic regression models with two dichotomous predictors can be found in table 2.

Table 1

Summary of Results and Conclusions for Simulated Logistic Regression Models with Continuous X_1 and Dichotomous X_2

Base Rate Effects	Measurement Reliability Effects	Sample Size Effects
1) At base rate = .10 there was a lot of index mean variation across cutoff points and across indices; 2) Index means became more consistent across cutoff points and across indices as the base rate approached .50; 3) Φ_p and τ_p consistently were closest in value; 4) Φ_p was the most base rate invariant, followed by τ_p, followed by λ_p; 5) Not much influence of base rate on RIOC nor percentage correct classifications.	1) Not much base rate influence on indices unless the continuous predictor was measured with low reliability; 2) As long as the continuous predictor had high reliability, it did not matter what the reliability level of the dichotomous predictor was.	Index means were similar across sample sizes, but the standard deviations were larger for the small sample scenario.

Table 2

Summary of Results and Conclusions for Simulated Logistic Regression Models with Two Dichotomous Predictors

Base Rate Effects	Measurement Reliability Effects	Sample Size Effects
1) There was a substantial base rate influence on all 3 predictive efficiency indices; 2)Index means were more consistent across cutoff points within a given base rate than across base rate levels; 3) At base rate = .10, Φ_p and λ_p were closest in mean value; but as the base rate approached .50 all index means became more consistent with each other; 4) T_p was most base rate invariant, followed by Φ_p, followed by λ_p; 5) RIOC often not computable; 6) Both RIOC and percentage correct classifications indices were more influenced by base rate changes than when a continuous predictor was included in the model.	The only time the reliability levels of the predictors were influential was when both predictors were measured with low reliability--had the effect of substantially lowering all index means (even when the model classified well).	Index means were similar across sample sizes, but the standard deviations were larger for the small sample scenario.

Recommendations

The recommendations for this study will be presented in two sections. The first section will discuss recommendations regarding the use of the predictive efficiency indices to assess the amount of improved predictive ability, over and above chance prediction, for logistic regression models. A table summarizing these recommendations will be presented as well. The second section will discuss recommendations for future research.

Use Of Predictive Efficiency Indices

Based on the conclusions of this current study, it is recommended that researchers utilize the ϕ_p index when estimating the predictive efficiency of their logistic regression models. Also, it should be kept in mind that if any continuously measured predictors are included in the logistic regression model, ϕ_p will be the most base rate invariant index. But if the model contains only dichotomous predictors, τ_p will be the least sensitive to base rate fluctuations.

Further, because researchers often conduct their research in a manner that involves some nonrandomized selection of subjects, they typically do not know the actual base rates of their samples. Since ϕ_p generally was the most base rate invariant index across the majority of model conditions simulated in this study, further support was provided regarding the recommendation to use the ϕ_p index.

While ϕ_p was found to be the most stable index across the various logistic regression model conditions, it is recommended that researchers compute all three of Menard's (1995) predictive efficiency indices. By comparing the values obtained across the three indices, some indication might be provided as to the underlying base rate of the sample. If all of the indices yield similar values, it could be inferred that the base rate is close to .50. On the other hand, if all of the indices yield distinctly disparate values, it could be inferred that the base rate is not close to .50, and the researcher then should use caution in interpreting the estimate of predictive efficiency. See Table 3 for a summary of these recommendations.

Table 3

Recommendations for Usage of Predictive Efficiency Indices

Model Conditions	Recommended Index
Base rate close to .50	It does not matter which of the 3 predictive efficiency indices are used; Φ_o was the most base rate invariant across all other model conditions.
Base rate smaller than .30; Model has Continuous X_1 and Dichotomous X_2	Use ϕ_o; Measure the continuous predictor with high reliability; RIOC can also be used if both predictors are not measured with low reliability.
Base rate smaller than .30; Model has Dichotomous X_1 and X_2	Use τ_o; RIOC often not calculable; RIOC can be used as long as both predictor variables are not measured with low reliability; While measurement reliability may have power implications for detecting statistically significant predictors, it did not appear to moderate base rate influences.
Large or Small Sample Size	It does not matter which of the indices to use; However, ϕ_o is recommended since it is the most base rate invariant index over the greatest number of model conditions; While sample size may have power implications for detecting statistically significant predictors, it did not appear to moderate base rate influences.
Uncertainty Regarding Base Rate	Use ϕ_o since it is the most base rate invariant index over the greatest number of model conditions.

Future Research

One recommendation for future research is that these same indices should continue to be investigated for base rate influence within other designs of logistic regression models. It is possible that there are other predictor variable combinations which alter the patterns of the index means.

BEST COPY AVAILABLE

Additionally, a future investigation of the influence of measurement reliability on predictive efficiency indices should include a component to explore these influences as they pertain to the criterion variable. The current study only investigated the influence of measurement reliability in the predictor variables. It would be interesting to determine if the patterns detected in this study would change if the reliability of the criterion variable was manipulated as well.

A final recommendation is that the search for a base rate invariant predictive efficiency index continues. As long as model conditions dictate the appropriateness of an index for assessing efficiency, ambiguity will continue to exist regarding which index to use, and when to use it.

References

Allen, M.J., & Yen, W.M. (1979). Introduction to Measurement Theory. Monterey, CA: Brooks/Cole.

Berends, M., Koretz, D., & Harris, E. (1995). Identifying Students at Risk of Low Achievement in NAEP and NELS. Washington, D.C.: National Center for Education Statistics. (ERIC Document Reproduction Service No. ED 404372).

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37-46.

Collis, G.M. (1985). Kappa, measures of marginal symmetry and intraclass correlations. Educational and Psychological Measurement, 45, 55-62.

Davis, J.A. (1971). Elementary Survey Analysis. Englewood Cliffs, NJ: Prentice-Hall.

Eimers, M.T. (1995). Exploring faculty career progression: A retention and tenure perspective. Paper presented at the Annual Forum of the Association for Institutional Research, 35th, Boston, MA, May 28-31. (ERIC Document Reproduction Service No. ED 386998).

Fergusson, D.M., Fifield, J.K., & Slater, S.W. (1977). Signal Detectability Theory and the Evaluation of Prediction Tables. Journal of Research in Crime and Delinquency, 14, 237-246.

French, A.W., & Miller, T.R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. Journal of Educational Measurement, 33(3), 315-352. (ERIC Document Reproduction Service No. EJ 535138).

Gillespie, M., & Noble, J. (1992). Factors Affecting Student Persistence: A Longitudinal Study. Iowa City, IA: American College Testing Program. (ERIC Document Reproduction Service No. ED 357056).

Gottfredson, D.M. & Tonry, M. (1987). Prediction and Classification: Criminal Justice Decision Making. Series: Crime and Justice, (Vol. 9). Chicago: University of Chicago Press.

Hosmer, D.W., & Lemeshow, S. (1989). Applied Logistic Regression. New York: Wiley.

Huesman, R. L., Moore, J.E., Huang, C.Y., & Guo, S. (1996). Identifying students at risk: Utilizing traditional and non-traditional data sources. Paper presented at the Annual Forum of the Association for Institutional Research, 36th, Albuquerque, NM, May 5-8. (ERIC Document Reproduction Service No. ED 397726).

Landis, J.R., & Koch, G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.

Menard, S. (1995). Applied Logistic Regression Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA: Sage.

Miller, D., Brownell, M.T., & Smith, S. (1995). Retention and attrition in special education: Analysis of variables that predict staying, transferring, or leaving. In the National Dissemination Forum on Issues Relating to Special Education Teacher Satisfaction, Retention and Attrition, Washington, DC, May 25-26. (ERIC Document Reproduction Service No. ED 389157).

Nunnally, J. (1978). Psychometric Theory, (2nd Ed.). New York: McGraw Hill.

Ryan, K.E., & Bachman, L.F. (1992). Differential Item Functioning on two tests of EFL proficiency. Language Testing, 9(1), 12-29. (ERIC Document Reproduction Service No. EJ 457604).

Ryan, K.E., & Chiu, S. (1996). Detecting DIF on mathematics items: The case for gender and calculator sensitivity. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY, April 8-12. (ERIC Document Reproduction Service No. ED 395998).

SAS Institute (1990). SAS/STAT User's Guide (Release 6.04). Cary, NC: Author.

Sherry, A.C., & Sherry, F.T. (1996). Computer confidence: Factors associated with retention in the community college. In Proceedings of Selected Research and Development Presentations at the 1996 National Convention of the Association for Educational Communications and Technology, 18th, Indianapolis, IN. (ERIC Document Reproduction Service No. ED 397839).

Soderstrom, I.R. (1997). Investigation of the base rate problem in predictive efficiency indices associated with logistic regression models. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.

Soderstrom, I.R., & Han, T. (1993). Effects of measurement error on the pretest covariate in three designs of experiments using analysis of covariance. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Soderstrom, I.R. & Leitner, D.W. (1996). Investigation of the base rate problem in predictive efficiency indices associated with logistic regression models. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Waltz, C.F., Strickland, O.L., & Lenz, E.R. (1991). Measurement in Nursing Research, (2nd Ed.). Philadelphia, PA: F.A. Davis Company.

Weimer, D. (1996). Applying linear and logistic regression to a required English proficiency test. Paper presented at the Annual Forum of the Association for Institutional Research, 36th, Albuquerque, NM, May 5-8. (ERIC Document Reproduction Service No. ED 397727).

Wilson, R.L., & Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. Educational and Psychological Measurement, 55(2), 186-195. (ERIC Document Reproduction Service No. EJ 505866).

BEST COPY AVAILABLE